

Package ‘nutshell’

January 27, 2015

Type Package

Title Data for ‘R in a Nutshell’

Version 2.0

Date 2012-12-04

Author Joseph Adler

Maintainer Joseph Adler <rinanutshell@gmail.com>

Description This package contains data sets used as examples in the book ‘R in a Nutshell’ from O’Reilly Media. For more information on this book, see <http://shop.oreilly.com/product/0636920022008.do>

License CC BY-NC-ND 3.0 US

LazyLoad yes

ZipData no

Depends R (>= 2.10), nutshell.bbdb, nutshell.audioscrobbler

Suggests MASS, RSQLite, arules, lattice, survival

Repository CRAN

Date/Publication 2012-12-12 09:14:03

NeedsCompilation no

R topics documented:

batting.2008	2
births2006.smpl	4
consumption	5
doctorates	6
dow30	7
field.goals	8
GSE2034	9
ham.price.ts	10
medicare.payments	11
medicare.payments.by.state	12

mort06.smpl	13
outcome.of.care.measures.national	14
sanfrancisco.home.sales	15
shiller	17
spambase	18
SPECint2006	20
team.batting.00to08	21
tires.sus	22
top.bacon.searching.cities	23
toxins.and.cancer	24
turkey.price.ts	25
yosemite	26

Index	27
--------------	-----------

batting.2008	<i>MLB Batting Data, 2008 Season</i>
--------------	--------------------------------------

Description

This data frame contains information on all 1384 players who had at least one plate appearance in MLB in 2008.

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Usage

```
data(batting.2008)
```

Format

A data frame with 1384 observations on the following 31 variables.

```
nameLast a character vector
nameFirst a character vector
weight a numeric vector
height a numeric vector
bats a character vector
throws a character vector
debut a character vector
birthYear a numeric vector
playerID a character vector
yearID a numeric vector
stint a numeric vector
teamID a character vector
```

lgID a character vector
G a numeric vector
G_batting a numeric vector
AB a numeric vector
R a numeric vector
H a numeric vector
2B a numeric vector
3B a numeric vector
HR a numeric vector
RBI a numeric vector
SB a numeric vector
CS a numeric vector
BB a numeric vector
SO a numeric vector
IBB a numeric vector
HBP a numeric vector
SH a numeric vector
SF a numeric vector
GIDP a numeric vector
G_old a numeric vector

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

This data is taken from the Baseball Databank database, available from <http://www.baseball-databank.org/>

Examples

```
data(batting.2008)  
tapply(X=batting.2008$HR, INDEX=list(batting.2008$teamID), FUN=sum)
```

`births2006.sml`*Births in the United States, 2006*

Description

This data set contains information on babies born in the United States during 2006. There is one record per birth. This data set is a random ten percent sample.

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media. Indices for observations from the original data set that are used in this sample are included in `births2006.idx`.

Usage

```
data(births2006.sml)
```

Format

A data frame with 427323 observations on the following 13 variables.

DOB_MM Month of date of birth

DOB_WK Day of week of birth

MAGER Mother's age

TBO_REC Total birth order

WTGAIN Weight gain by mother

SEX a factor with levels F M, representing the sex of the child

APGAR5 APGAR score

DMEDUC Mother's education level

UPREVIS Number of prenatal visits

ESTGEST Estimated weeks of gestation

DMETH_REC Delivery Method

DPLURAL "Plural Births;" levels include 1 Single, 2 Twin, 3 Triplet, 4 Quadruplet, and 5 Quintuplet or higher

DBWT Birth weight, in grams

Source

http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm

Examples

```
data(births2006.sml)
library(lattice)
histogram(~DBWT|DPLURAL,data=births2006.sml)
densityplot(~DBWT,groups=DPLURAL,data=births2006.sml,
            plot.points=FALSE,auto.key=TRUE)
```

`consumption`*Per capita US Food Consumption 1980-2005*

Description

This data frame contains information on the per-capita consumption of food in the United States, between 1980 and 2005. The data is taken from the Statistical Abstract of the United States.

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Usage

```
data(consumption)
```

Format

A data frame with 42 observations on the following 4 variables.

`Food` a factor with levels Caloric sweeteners, Dairy products, Eggs, Fish and Shellfish, Flour and cereal products, Poultry, Red Meat

`Units` a factor with levels Number, Pounds

`Year` a numeric vector

`Amount` a numeric vector

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

http://www.census.gov/compendia/statab/cats/health_nutrition/food_consumption_and_nutrition.html

Examples

```
data(consumption)
library(lattice)
dotplot(
  Amount ~ Year | Food,
  data=consumption,
  aspect="xy",
  scales=list(relation="sliced", cex=.4)
)
```

doctorates

Doctorates Awarded in the United States, 2001-2006

Description

This data set contains information on the number of doctorate degrees awarded in the United States between 2001 and 2006.

Usage

```
data(doctorates)
```

Format

A data frame with 6 observations on the following 7 variables.

year a numeric vector

engineering a numeric vector

science a numeric vector

education a numeric vector

health a numeric vector

humanities a numeric vector

other a numeric vector

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

This data is from the Statistical Abstract of the United States. You can download the data from <http://www.census.gov/compendia/statab/tables/09s0785.xls>

Examples

```
data(doctorates)
# make this into a matrix:
doctorates.m <- as.matrix(doctorates[2:7])
rownames(doctorates.m) > doctorates.m
barplot(doctorates.m, beside=TRUE, horiz=TRUE, legend=TRUE, cex.names=.75)
barplot(t(doctorates.m), legend=TRUE, ylim=c(0, 66000))
```

dow30	<i>Dow 30 Stock Quotes This data frame contains stock quotes for all 30 stocks in the Dow Jones Industrial Average between September 22, 2008 and September 21, 2009.</i>
-------	---

Description

Dow 30 Stock Quotes

Usage

```
data(dow30)
```

Format

A data frame with 7482 observations on the following 8 variables.

symbol A factor representing the date for each quote

Date Date on which the quote was taken

Open A numeric vector representing the opening price for each stock on each day

High A numeric vector representing the high price for each stock on each day

Low A numeric vector representing the low price for each stock on each day

Close A numeric vector representing the closing price for each stock on each day

Volume A numeric vector representing the number of shares traded for each stock on each day

Adj.Close A numeric vector representing the closing price for each stock on each day

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

The data in this data frame was obtained from Yahoo! Finance, <http://finance.yahoo.com/>.

Examples

```
data(dow30)
summary(dow30)
```

`field.goals`*2005 Field Goal Attempts*

Description

This data set contains information on all 982 field goal attempts during the 2005 NFL season.

Usage

```
data(field.goals)
```

Format

A data frame with 982 observations on the following 10 variables.

`home.team` a factor representing the home team

`week` a numeric vector representing the week of the season

`qtr` a numeric vector representing the quarter during the game

`away.team` a factor representing the away team

`offense` a factor representing the offensive team

`defense` a factor representing defensive team

`play.type` a factor with levels FG aborted, FG blocked, FG good, and FG no

`player` a factor representing player names

`yards` a numeric vector

`stadium.type` a factor with levels Both, In, and Out

Details

This data is provided courtesy of Aaron Schatz, Editor-in-Chief of Football Outsiders a web site about American Football.

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

<http://www.footballoutsiders.com>

Examples

```
data(field.goals)
hist(field.goals$yards)
field.goals.forlr <- transform(field.goals,
  good=as.factor(ifelse(play.type=="FG good", "good", "bad")))
field.goals.table <- table(field.goals.forlr$good,
  field.goals.forlr$yards)
plot(colnames(field.goals.table),
```



```

    field.goals.table["good",] /
    (field.goals.table["bad",] +
     field.goals.table["good",]),
    xlab="Distance (Yards)",
    ylab="Percent Good"
  )

```

GSE2034

Breast cancer relapse free survival

Description

This data set represents 180 lymph-node negative relapse free patients and 106 lymph-node negative patients that developed a distant metastasis. It contains information on the time until relapse of last follow up for each patient, an indicator of whether the ER gene was expressed, and an indicator whether a relapse occurred.

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Usage

```
data(GSE2034)
```

Format

A data frame with 286 observations on the following 7 variables.

PID a numeric vector

GEO.accession.number The unique identifier for each subject on GEO

lymph.node.status a factor with levels negative

months.to.relapse.or.last.followup a numeric vector

relapse a logical vector

ER.Status a factor with levels ER- and ER+

Brain.relapses a logical vector

Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2034>

Examples

```

data(GSE2034)
library(survival)
GSE2034.Surv <- transform(GSE2034,
  surv=Surv(
    time=GSE2034$months.to.relapse.or.last.followup,
    event=GSE2034$relapse,
    type="right"
  )
)

```

```
)  
)  
GSE2034.survfit <- survfit(  
  formula=surv~ER.Status,  
  data=GSE2034.Surv )  
plot(GSE2034.survfit,lty=1:2)  
legend(135,1,c("ER+", "ER-"),lty=1:2,cex=0.5)
```

ham.price.ts

Ham Price Time Series

Description

A time series objects consisting of average monthly retail prices per pound of ham in the United States between January 2001 and April 2008.

Usage

```
data(ham.price.ts)
```

Format

The format is: Time-Series [1:88] from January 2001 to April 2008

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

This data is from the Livestock Market Information Center (LMIC) Retail Scanner Prices for Meat database. For more information on this data source, see <http://www.lmic.info/meatscanner/meatscanner.shtml>

Examples

```
data(ham.price.ts)  
data(turkey.price.ts)  
ccf(turkey.price.ts, ham.price.ts, plot=FALSE)
```

medicare.payments	<i>Average Medicare Payment Data</i>
-------------------	--------------------------------------

Description

A data set that contains average medicare payments by hospital for 70 common medical conditions. This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Usage

```
data(medicare.payments)
```

Format

A data frame with 140722 observations on the following 14 variables.

Provider.Number A unique ID for the provider

Hospital.Name The name of the care provider

Address.1 First line of the address for the care provider

Address.2 Second line of the address for the care provider

Address.3 Third line of the address for the care provider

City City in which the care provider is located

State The state in which the care provider is located

ZIP.Code The zip code of the care provider

County.Name The county in which the care provider is located

Phone.Number The phone number for the care providers

Diagnosis.Related.Group A description of the medical condition

Medicare.Average.Payment a numeric vector that shows the average medicare repayment for the specified condition and hospital

Number.Of.Cases a numeric vector that specifies the number of cases

Footnote a factor with levels An asterisk (*) appears in the table where data cannot be disclosed to protect

Source

http://www.medicare.gov/download/Hospital_Revised_flatfiles.zip

Examples

```
data(medicare.payments)
```

```
medicare.payments.by.state
```

Medicare Payments by State

Description

A data set that contains average medicare payments by state for 70 common medical conditions.

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Usage

```
data(medicare.payments.by.state)
```

Format

A data frame with 3735 observations on the following 6 variables.

`State` A factor that specifies the state for which this statistic is calculated

`Diagnosis.Related.Group` A description of the medical condition

`Medicare.Average.Payment.Minimum` The minimum average hospital medicare payment for this condition and state

`Medicare.Average.Payment.Maximum` The maximum average hospital medicare payment for this condition and state

`Number.Of.Cases` The number of cases seen; An asterisk (*) appears in the table where data cannot be disclosed to protect personal health information due to the small number of Medicare patients (fewer than 11)

`Footnote` A footnote from the original data

Source

http://www.medicare.gov/download/Hospital_Revised_flatfiles.zip

Examples

```
data(medicare.payments.by.state)
```

`mort06.smpl`*US Mortality Data, 2006*

Description

A 10

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Usage

```
data(mort06.smpl)
```

Format

A data frame with 243073 observations on the following 35 variables.

ResidentStatus a numeric vector

Education1989 a numeric vector, Education (1989 revision)

Education2003 a numeric vector, Education (2003 revision)

EducationFlag a numeric vector, Education reporting flag

MonthOfDeath a numeric vector, Month of Death

Sex a factor with levels F and M

AgeDetail a numeric vector, Detail Age

AgeSubstitution a numeric vector, Age Substitution Flag

AgeRecode52 a numeric vector

AgeRecode27 a numeric vector

AgeRecode12 a numeric vector

AgeRecodeInfant22 a numeric vector

PlaceOfDeath a numeric vector

MaritalStatus a factor with levels D, M, S, U, and W

DayOfWeekofDeath a numeric vector

CurrentDataYear a numeric vector

InjuryAtWork a factor with levels N, U, and Y

MannerOfDeath a numeric vector

MethodOfDisposition a factor with levels B, C, D, E, O, R, and U

Autopsy a factor with levels N, U, and Y

ActivityCode a numeric vector

PlaceOfInjury a numeric vector

ICDCode a factor

Cause a factor describing the cause of death

CauseRecode358 a numeric vector, 358 Cause Recode
 CauseRecode113 a numeric vector, 113 Cause Recode
 CauseRecode130 a numeric vector, 130 Infant Cause Recode
 CauseRecord39 a numeric vector, 39 Cause Recode
 Race a factor with levels American Indian (includes Aleuts and Eskimos) Asian Indian
 Black Chinese Combined other Asian or Pacific Islander Filipino Guamanian
 Hawaiian (includes Part-Hawaiian) Japanese Korean Other Asian or Pacific Islander
 Samoan Vietnamese White
 BridgeRaceFlag a numeric vector
 RaceImputationFlag a numeric vector
 RaceRecode3 a numeric vector
 RaceRecord5 a numeric vector
 HispanicOrigin a numeric vector
 HispanicOriginRecode a numeric vector
 age a numeric vector

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm

Examples

```
data(mort06.smpl)
aov(age~Cause,data=mort06.smpl)
model.tables(aov(age~Cause, data=mort06.smpl))
```

outcome.of.care.measures.national

National outcome of care measures

Description

A small data set that shows the national average mortality and readmission rates for heart attacks, heart failure, and pneumonia

Usage

```
data(outcome.of.care.measures.national)
```

Format

A data frame with 6 observations on the following 3 variables.

Condition a factor that describes the statistics: Heart Attack, Heart Failure, or Pneumonia

Measure a factor that describes the measure: Heart Attack Death (Mortality) Rates, Heart Attack Readmission Rates, Heart Failure Death (Mortality) Rates, Heart Failure Readmission Rates, Pneumonia (PN) 30-Day Mortality Rate, Pneumonia (PN) 30-Day Readmission Rates, Mortality, or Readmission

Rate the quantity being measured

Source

http://www.medicare.gov/download/Hospital_Revised_flatfiles.zip

Examples

```
data(outcome.of.care.measures.national)
```

```
sanfrancisco.home.sales
```

San Francisco Home Sales Data

Description

This data contains information on homes sold in San Francisco between 2/13/2008 and 7/14/2009.

Usage

```
data(sanfrancisco.home.sales)
```

Format

A data frame with 3281 observations on the following 15 variables.

line a numeric vector representing the line number of the observation in the data set

county a factor with levels San Francisco County

street a factor representing the street address of the property

city a factor with levels San Francisco

zip a numeric vector representing the zip code of the property

date a Date representing the sale date

price a numeric vector representing the sales price

bedrooms a numeric vector representing the number of bedrooms

squarefeet a numeric vector representing the interior are of the property, in square feet

lotsize a numeric vector representing the lot size of the property, in square feet

year a numeric vector representing the year in which the property was built
 latitude a numeric vector representing the latitude coordinate of the property
 longitude a numeric vector representing the longitude coordinate of the property
 month a factor representing the month in which the property was sold
 neighborhood a factor representing neighborhood names

Details

This data set was assembled from a variety of sources, including two Bay area newspapers (the San Jose Mercury News and the San Francisco Chronicle), Yahoo Maps, and Zillow Neighborhood Boundaries.

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media. In the book, we took separate samples for training and testing. Indices for observations in each sample are included in `sanfrancisco.home.sales.testing.indices` and `sanfrancisco.home.sales.training.indices`.

Source

Data was assembled from a variety of sources including <http://www.sfgate.com> <http://www.mercurynews.com> <http://www.zillow.com/howto/api/neighborhood-boundaries.htm>

Examples

```

data(sanfrancisco.home.sales)
library(lattice)
trellis.par.set(fontsize=list(text=7))
dollars.per.squarefoot <- mean(
  sanfrancisco.home.sales$price / sanfrancisco.home.sales$squarefeet,
  na.rm=TRUE);
xyplot(price~squarefeet|neighborhood,
  data=sanfrancisco.home.sales,
  pch=19,
  cex=.2,
  subset=(zip!=94100 & zip!=94104 & zip!=94108 &
    zip!=94111 & zip!=94133 & zip!=94158 &
    price<4000000 &
    ifelse(is.na(squarefeet),FALSE,squarefeet<6000)),
  strip=strip.custom(strip.levels=TRUE,
    horizontal=TRUE,
    par.strip.text=list(cex=.8)),
  panel=function(...) {
    panel.abline(a=0,b=dollars.per.squarefoot);
    panel.xyplot(...);
  }
)

```

`shiller`*Shiller Home Price Index*

Description

Robert J Shiller is an economics professor at Yale University, and one of the leading experts on housing prices in the United States. This data set contains his home price index (and some other data) over the past century.

Usage

```
data(shiller)
```

Format

A data frame with 126 observations on the following 2 variables.

Year a numeric vector

Real.Home.Price.Index a numeric vector

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Other information (including long bond rates, US population size, and cost of construction) is included in `shiller.other.data`.

Source

<http://www.irrationalexuberance.com/>

Examples

```
data(shiller)
# loads shiller.index, shiller.other.data
# linear fit
hpi.lm <- lm(Real.Home.Price.Index~Year,data=shiller.index)
# plotting the fit
plot(shiller.index,pch=19,cex=0.3)
abline(reg=hpi.lm,lty=1)
```

spambase

Spambase Data Set

Description

The Spambase data set was created by Mark Hopkins, Erik Reeber, George Forman, and Jaap Suermondt at Hewlett-Packard Labs. It includes 4601 observations corresponding to email messages, 1813 of which are spam. From the original email messages, 58 different attributes were computed.

Usage

```
data(spambase)
```

Format

A data frame with 4601 observations on the following 58 variables.

```
word_freq_make a numeric vector  
word_freq_address a numeric vector  
word_freq_all a numeric vector  
word_freq_3d a numeric vector  
word_freq_our a numeric vector  
word_freq_over a numeric vector  
word_freq_remove a numeric vector  
word_freq_internet a numeric vector  
word_freq_order a numeric vector  
word_freq_mail a numeric vector  
word_freq_receive a numeric vector  
word_freq_will a numeric vector  
word_freq_people a numeric vector  
word_freq_report a numeric vector  
word_freq_addresses a numeric vector  
word_freq_free a numeric vector  
word_freq_business a numeric vector  
word_freq_email a numeric vector  
word_freq_you a numeric vector  
word_freq_credit a numeric vector  
word_freq_your a numeric vector  
word_freq_font a numeric vector  
word_freq_000 a numeric vector
```

word_freq_money a numeric vector
word_freq_hp a numeric vector
word_freq_hpl a numeric vector
word_freq_george a numeric vector
word_freq_650 a numeric vector
word_freq_lab a numeric vector
word_freq_labs a numeric vector
word_freq_telnet a numeric vector
word_freq_857 a numeric vector
word_freq_data a numeric vector
word_freq_415 a numeric vector
word_freq_85 a numeric vector
word_freq_technology a numeric vector
word_freq_1999 a numeric vector
word_freq_parts a numeric vector
word_freq_pm a numeric vector
word_freq_direct a numeric vector
word_freq_cs a numeric vector
word_freq_meeting a numeric vector
word_freq_original a numeric vector
word_freq_project a numeric vector
word_freq_re a numeric vector
word_freq_edu a numeric vector
word_freq_table a numeric vector
word_freq_conference a numeric vector
char_freq_semicolon a numeric vector
char_freq_left_paren a numeric vector
char_freq_left_bracket a numeric vector
char_freq_exclamation a numeric vector
char_freq_dollar a numeric vector
char_freq_pound a numeric vector
capital_run_length_average a numeric vector
capital_run_length_longest a numeric vector
capital_run_length_total a numeric vector
is_spam a factor with levels 0 1

Details

This data is used as an example in the book "R in a Nutshell," from O'Reilly Media.

Source

This data set is from the UCI Machine Learning Repository. You can find more information about this data set, including the citation policy, from <http://archive.ics.uci.edu/ml/datasets/Spambase>

Examples

```
data(spambase)
table(spambase$is_spam)
# fit a linear discriminant analysis model to the data
library(MASS)
spam.lda <- qda(formula=is_spam~., data=spambase)
```

SPECint2006

SPECint2006 Results

Description

This data set contains results from the SPECint2006 benchmarks.

Usage

```
data(SPECint2006)
```

Format

A data frame with 1233 observations on the following 9 variables.

Benchmark a factor with levels CINT2006

Hardware.Vendor a factor representing the name of the vendor

System a factor representing the name of the system

Num.Cores a numeric vector representing the number of cores on the system

Num.Chips a numeric vector representing the number of CPUs on the system

Num.Cores.Per.Chip a numeric vector

Result a numeric vector representing the benchline result

Baseline a numeric vector representing the baseline result

Published a factor representing the month in which the benchmark was published

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

<http://www.spec.org/cpu2006/>

Examples

```
data(SPECint2006)
t.test(subset(SPECint2006,Num.Chips==1&Num.Cores==2)$Baseline,
       subset(SPECint2006,Num.Chips==1&Num.Cores==2)$Result,
       paired=TRUE)
```

team.batting.00to08 *MLB Team Batting Statistics, 2000-2008*

Description

This data set contains statistics on team batting between 2000 and 2008.

Usage

```
data(team.batting.00to08)
```

Format

A data frame with 270 observations on the following 13 variables.

teamID a character vector
yearID a numeric vector
runs a numeric vector
singles a numeric vector
doubles a numeric vector
triples a numeric vector
homeruns a numeric vector
walks a numeric vector
stolenbases a numeric vector
caughtstealing a numeric vector
hitbypitch a numeric vector
sacrificeflies a numeric vector
atbats a numeric vector

Details

This data contains aggregate offensive statistics for each team in Major League Baseball between 2000 and 2008. It is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

The data was derived from the Baseball Databank database, downloadable from <http://www.baseball-databank.org>.

Examples

```

data(team.batting.00to08)
attach(team.batting.00to08)
runs.mdl <- lm(
  formula=runs~singles+doubles+triples+homeruns+
    walks+hitbypitch+sacrificeflies+
    stolenbases+caughtstealing,
  data=team.batting.00to08)
summary(runs.mdl)

```

tires.sus

Stepped Up Speed Tire Failure Test Data

Description

This data set contains the results of an experiment on tire durability conducted by the NHTSA.

Usage

```
data(tires.sus)
```

Format

A data frame with 66 observations on the following 27 variables.

Phase a numeric vector

Tire_Type a factor with levels B, C, D, E, H, and L

Barcode a numeric vector

Dot_Number a factor

Dot_MidWeekDate a factor

Collection_Date a factor

DOT_Age a numeric vector

X1st_Task a factor with levels HS_AIR ,HS_AIR_F, and HS_AIR_F_S

X1st_Task_Status a factor with levels Tested

Position a factor with levels Left Front, Left Rear, New, Right Front, Right Rear, and Spare

ORN a factor with levels New, OEM, and Repl

AZ_Use a numeric vector

DOT_Est_Mileage_mi a numeric vector

DOT_Est_Mileage_km a numeric vector

Initial_IP_kPa a numeric vector

Load_kg a numeric vector

Time_To_Failure a numeric vector

Speed_At_Failure_km_h a numeric vector
Mileage_At_Failure_km a numeric vector
Millions_Cycles_At_Failure a numeric vector
Failure_Type a factor
Failure.Notes a factor
Photo_1 a factor
Photo_2 a factor
Photo_3 a factor
Photo_4 a factor
Invoice_Date a factor representing invoice dates

Details

This data is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

The data for this example was taken from <http://www-nrd.nhtsa.dot.gov/vrtc/ca/tires.htm>

Examples

```
data(tires.sus)
library(lattice)
dotplot(as.factor(Speed_At_Failure_km_h) ~ Time_To_Failure | Tire_Type,
        data=tires.sus)
```

top.bacon.searching.cities
Top Bacon Searching Cities

Description

This data ranks US cities by the frequency with which residents of those cities search for the word "Bacon" on Google.

Usage

```
data(top.bacon.searching.cities)
```

Format

A data frame with 15 observations on the following 2 variables.

city a factor with levels representing US city names
rank a numeric vector

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

The data was from Google Insights, <http://www.google.com/insights/search/#q=bacon&cmpt=q>. Query was run on September 5 2009, for data from 2004 through 2009.

Examples

```
data(top.bacon.searching.cities)
typeof(top.bacon.searching.cities)
class(top.bacon.searching.cities)
```

toxins.and.cancer	<i>Toxins and Cancer</i>
-------------------	--------------------------

Description

This data set contains information on the volume of toxic chemicals released in each state during 2006, deaths from cancer during 2008, and the surface area of each state.

Usage

```
data(toxins.and.cancer)
```

Format

A data frame with 41 observations on the following 15 variables.

State a factor with levels Alabama Alaska Arizona ...

total_toxic_chemicals a numeric vector

total_on_site a numeric vector

air_on_site a numeric vector

other_on_site a numeric vector

off_site a numeric vector

Surface_Area a numeric vector

new_total a numeric vector

new_breast a numeric vector

new_lung a numeric vector

deaths_total a numeric vector

deaths_breast a numeric vector

deaths_lung a numeric vector

Population a numeric vector

State_Abbrev a factor with levels AK, AL, AR ...

Details

This data is used as an example in the book "R in a Nutshell," from O'Reilly Media.

Source

This data was taken from several tables in the Statistical Abstract of the United States. You can download this data from <http://www.census.gov/compendia/statab>

Examples

```
data(toxins.and.cancer)
attach(toxins.and.cancer)
plot(total_toxic_chemicals/Surface_Area,deaths_total/Population)
plot(air_on_site/Surface_Area,deaths_lung/Population)
```

turkey.price.ts

Monthly Average Turkey Price, January 2001 to April 2008

Description

This time series shows the average retail price of turkey in the United States between January 2001 and April 2008.

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Usage

```
data(turkey.price.ts)
```

Format

The format is: Time-Series [1:88] from 2001 to 2008: 1.58 1.75 1.63 1.45 1.56 2.07 1.81 1.74 1.54 1.45 ...

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media.

Source

This data is from the Livestock Market Information Center (LMIC) Retail Scanner Prices for Meat database. For more information on this data source, see <http://www.lmic.info/meatscanner/meatscanner.shtml>

Examples

```
data(turkey.price.ts)
start(turkey.price.ts)
end(turkey.price.ts)
frequency(turkey.price.ts)
plot(turkey.price.ts)
```

`yosemite`*Yosemite Valley Elevation Data*

Description

This matrix contains information on the elevation at different points in Yosemite Valley. It is useful as an example for plotting 3 dimensional data.

Usage

```
data(yosemite)
```

Format

The format is: num [1:562, 1:253] 2351 2345 2338 2331 2322 ...

Details

This data set is used as an example in the book "R in a Nutshell" from O'Reilly Media. You can find directions on how to create a similar data set for another location in the book.

Source

This data was taken from the US Geological Survey's National Map Seamless Server: <http://seamless.usgs.gov/website/seaml>

Examples

```
data(yosemite)
yosemite.flipped <- yosemite[,seq(from=253,to=1)]
yosemite.rightmost <- yosemite[nrow(yosemite) - ncol(yosemite) + 1,]
halfdome <- yosemite[(nrow(yosemite) - ncol(yosemite) + 1):562,
                    seq(from=253,to=1)]
persp(halfdome, col=grey(.25), border=NA, expand=.15,
      theta=225, phi=20, ltheta=45, lphi=20, shade=.75)
```

Index

*Topic **datasets**

- batting.2008, [2](#)
 - births2006.smpl, [4](#)
 - consumption, [5](#)
 - doctorates, [6](#)
 - dow30, [7](#)
 - field.goals, [8](#)
 - GSE2034, [9](#)
 - ham.price.ts, [10](#)
 - medicare.payments, [11](#)
 - medicare.payments.by.state, [12](#)
 - mort06.smpl, [13](#)
 - outcome.of.care.measures.national, [14](#)
 - sanfrancisco.home.sales, [15](#)
 - shiller, [17](#)
 - spambase, [18](#)
 - SPECint2006, [20](#)
 - team.batting.00to08, [21](#)
 - tires.sus, [22](#)
 - top.bacon.searching.cities, [23](#)
 - toxins.and.cancer, [24](#)
 - turkey.price.ts, [25](#)
 - yosemite, [26](#)
-
- batting.2008, [2](#)
 - births2006.idx (births2006.smpl), [4](#)
 - births2006.smpl, [4](#)
-
- consumption, [5](#)
-
- doctorates, [6](#)
 - dow30, [7](#)
-
- field.goals, [8](#)
-
- GSE2034, [9](#)
-
- ham.price.ts, [10](#)
-
- medicare.payments, [11](#)
-
- medicare.payments.by.state, [12](#)
 - mort06.smpl, [13](#)
 - outcome.of.care.measures.national, [14](#)
 - sanfrancisco.home.sales, [15](#)
 - shiller, [17](#)
 - spambase, [18](#)
 - SPECint2006, [20](#)
 - team.batting.00to08, [21](#)
 - tires.sus, [22](#)
 - top.bacon.searching.cities, [23](#)
 - toxins.and.cancer, [24](#)
 - turkey.price.ts, [25](#)
 - yosemite, [26](#)